ANALYSIS OF PATTERNS AMONG OBJECTS OF A PLURALITY OF CLASSES

Technical Field

This invention relates to a method for the automatic or interactive analysis of patterns among objects of a plurality of classes.

5 ## *Background of the Invention*

New advances in fluorescent probe technologies and imaging techniques are enabling the simultaneous imaging of many more biological entities (classes) than has been possible before. For example, spectral imaging (CRI Inc. http://www.cri-inc.com/instruments/products/index.shtml) , or emission finger-printing (Carl Zeiss LSM

10 510 META. http://www.zeiss.de/C12567BE0045ACF1/Inhalt-Frame/C98CD5EF1EFAF4EEC1256AC5003148E9) acquire images at multiple wavelengths and can generate precise optical spectra at every pixel. In a second step, this information is used for the digital separation of up to eight fluorophores. Spectral karyotyping (SKY) (Schrock E, du Manoir S, Veldman T, Schoell B, Wienberg J,

15 Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-Am I, Soenksen D, Garini Y, Ried T. 1996. Multicolor spectral karyotyping of human chromosomes. Science. 26;273(5274):494-7) and multiplexed fluorescence in situ hybridization (MFISH) (Speicher MR, Gwyn Ballard S, Ward DC. 1996. Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nat Genet. 12(4):368-75.) enable the simultaneous

20 visualization of the endogenous arrangement of the complete karyotype. These innovations increase the amount of biology we can resolve in an image by a factor of ten. Though the technologies underlying these innovations differ, the implications for biological image analysis are the same; current and next generation biological images will contain many more biological objects and relations, and many more classes of

25 objects than has been the case until now. There will be corresponding growth in the need among scientists in basic research, drug discovery and diagnostic imaging for a versatile and flexible tool to assist users for the detection and analysis of patterns in the relational, spectral, temporal arrangement of these individual objects and object classes.

30   Previously limited to three (usually one channel for the nuclear background, and a
      channel each for two chromosome classes), cytogeneticists can now see an entire
      karyotype simultaneously in a single SKY or MFISH image. Figures 1 and 2 demonstrate
      this profound advance. Figure 1 shows a typical standard FISH image of the interphase
      nucleus. There are three fluorescent images, Figure 1A, Figure 1B and Figure 1C. Object

35   mask images labeled by class can be produced from each channel image through image
      processing. See Figure 1D, Figure 1E and Figure 1F. Here there are three object classes;
      the nuclear envelope class containing one individual object (Figure 1A and Figure 1D), a
      chromosome class containing two individual chromosomes (homologs), Figure 1B and
      Figure 1E, and a second, different chromosome class containing two individual homologs

40   (Figure 1C and Figure 1F).

      In contrast, Figure 2 shows a SKY image of the metaphase rosette. SKY makes use of an
      interferogram to process multiple spectra at the pixel level, and algorithmically combines
      that information with a CCD image (Figure 2A) to produce an output image composed of

45   N individually segmented object masks (Figure 2B), where N corresponds to roughly two
      times the number of chromosomes in the karyotype (Schrock E, du Manoir S, Veldman
      T, Schoell B, Wienberg J, Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-Am I,
      Soenksen D, Garini Y, Ried T. 1996. Multicolor spectral karyotyping of human
      chromosomes. Science. 26;273(5274):494-7; Applied Spectral Imaging.

50   http://www.spectral-imaging.com). In Figure 2A, N=40 for these mouse cells (humans
      would have 46). Each mask is labeled via a look up table with a user defined RGB value
      corresponding to the object (chromosome) class. In Figure 2B there are two individual
      objects per object class.

55   Despite these advances, little software has been developed for the analysis of relational
      patterns among biological objects. The majority of today's analysis software is designed
      to measure the response of a biological system to perturbation (Giuliano K, Kapur R.
      2002. System for cell based screening. United States patent number 6,416,959; Harris
      TD, Hansen RL, Karsh W, Nicklaus NA, Trautman JK. 2002. Method and apparatus for

60   screening chemical compounds. United States patent number 6,388,788.), or to manually

or automatically score disease samples (TriPath Imaging Inc., FocalPoint Slide Profiler. http://www.tripathimaging.com/usproducts/focalpoint.htm; Applied Imaging Corp., CytoVision. http://www.appliedimagingcorp.com/usa/cyto/index.htm.) by analyzing the fluorescently labeled biology in the image. In stark contrast to this type of evaluation of

65   phenotypic characteristics of biological objects as an indicator in modern biology, software can assist human to analyze relational patterns in the location or relational, temporal arrangement of biological objects is virtually inexistence in the life sciences applications.

70   Due to the tedious nature of manual analysis as well as the lack of automatic analysis technology, scientists are using application-specific image analysis methods that generate only a very few patterns and few samples for analysis. The inability to rapidly create and analyze a large number of relational patterns makes it inefficient to find important characteristics. The domain of possible relational patterns and combinations of patterns is

75   large, and even assuming that one pattern will reveal a preference of relational arrangement (which indeed may not be the case), the possibility that any one pattern will reveal that preference is low. Furthermore, current methods suffer from low repeatability because they are based on imprecise image processing techniques developed for specific applications. These methods fail in the face of typical variations found in biological

80   images such as large variations in the shape, orientation and size of biological objects (even among the same class), image variations resulting from operator or equipment variability, and variation in image orientation.

It would be ideal for scientists to possess an efficient, robust, accurate and flexible tool

85   for pattern creation and review. Such a tool would enable them to distinguish the difference between a pattern signal and noise, and quickly find interesting relational patterns in biological images. The pattern sets could be scientist-generated rather than computer generated, this allows the analysis outcome to be easily validated.

90   Scientists have long been interested in the relational arrangement of biological objects. The idea that chromosomes may be arranged in a specific fashion has been considered

since the time of Boveri (Baltzer F. 1964. Theodor Boveri. Science. 15(144):809-15). A well-known demonstration of general chromosomal organization is the Rabl orientation, a polarization of centromeres and telemeres, observed in the early embryo of the fruit fly.

95   Patterns in the relational arrangement of biological objects have been studied outside the nucleus as well in skin cancer diagnosis (BC Cancer Research Centre. Research Arm of the BC Cancer Agency. http://www.bccrc.ca/ci/ta01_archlevel.html), retinal cell arrangement (Eglen SJ, Raven MA, Tamrazian E, Reese BE. 2003. Dopaminergic amacrine cells in the inner nuclear layer and ganglion cell layer comprise a single

100   functional retinal mosaic. J Comp Neurol. 446(3):343-55.), fungal spores arrangement (Jones CL, Lonergan GT, Mainwaring DE. Minimal spanning tree analysis of fungal spore relational patterns. Published online at http://www.swin.edu.au/chem/bio/fractals/mst01.htm) and platelets in wound healing (Beals M, Gross L, Harrell S. 2000. Cell aggregation and sphere packing. The Institute

105   for Environmental Modeling at the University of Tennessee).

Prior art detection and analysis of relational arrangement patterns is a two step process that requires image processing software, such as Universal Imaging's Metamorph (Universal Imaging Corp. Metamorph. http://www.image1.com/products/metamorph/) or

110   Media Cybernetics' ImagePro (Media Cybernetics Inc. ImagePro. http://www.mediacy.com/famip.htm) , to perform image segmentation and generate pattern measurements, and a data mining or statistics package such Spotfire's DecisionSite (Ahlberg C, Truve S, Wistrand E. 2000. System and method for automatic analysis of data bases and for user controlled dynamic querying. United States patent

115   number 6,014,661; Spotfire Inc. DecisionSite. http://www.spotfire.com/products/decision.asp) or Insightful's S-PLUS (Insightful S-PLUS. http://www.insightful.com/products/splus/default.asp) to evaluate the complex pattern signal. Even a single pattern signal can be complex because of the large number of object interactions. For example analyzing a nearest neighbor pattern across the

120   objects in Figure 2 would require the analysis of 1,560 interactions (the nearest neighbor pattern for relation i x j is not equivalent for j x i). Since there is no integrated approach,

it is difficult for scientists to iterate quickly between pattern creation and pattern analysis. This approach is far from general purpose, and inefficient for finding patterns.

125   Some examples of this type of prior art approach:

Eglen et al. (Eglen SJ, Raven MA, Tamrazian E, Reese BE. 2003. Dopaminergic amacrine cells in the inner nuclear layer and ganglion cell layer comprise a single functional retinal mosaic. J Comp Neurol. 446(3):343-55.) used Voronoi domains to

130   analyze the relational arrangement of dopaminergic amacrine cells in two different layers of the ferret retina to determine if the arrangement was different between layers, which might indicate distinct cell function for the cells in the different layers. Eglen calculated a regularity index statistic for each cell as the Voronoi domain area divided by the standard deviation of all Voronoi domain areas in the image.

135

The BC Cancer Research Centre uses Minimum Spanning Tree (MST) vertex length as a feature on which statistical analysis is performed (BC Cancer Research Centre. Research Arm of the BC Cancer Agency. http://www.bccrc.ca/ci/ta01_archlevel.html). These outputs are then linearly combined into an index that can be used for the quantitative

140   diagnosis of pre-neoplastic lesions. Jones et al. (Jones CL, Lonergan GT, Mainwaring DE. Minimal spanning tree analysis of fungal spore relational patterns. Published online at http://www.swin.edu.au/chem/bio/fractals/mst01.htm)also used MST to look at fungal spore relational arrangement and its relation to the asexual reproduction mechanism. Research at the Institute of Environmental Modeling, University of Tennessee has

145   utilized sphere packing theory to quantify the density of aggregated cells involved in wound healing (Beals M, Gross L, Harrell S. 2000. Cell aggregation and sphere packing. The Institute for Environmental Modeling at the University of Tennessee.)

Basic research in chromosome arrangement in the field of cytogenetics has been

150   rejuvenated and accelerated with recent discoveries that link the developmental regulation of lineage-restricted genes with their nuclear compartmentalization (Brown KE, Guest SS, Smale ST, Hahm K, Merkenschlager M, Fisher AG. 1997. Association of

transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. Cell. 91(6):845-54; Brown KE, Baxter J, Graf D, Merkenschlager M, Fisher AG. 1999.

155   Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. Mol Cell. 3(2):207-17; Kosak ST, Skok JA, Medina KL, Riblet R, Le Beu MM, Fisher AG, Sing H. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science. 296(5565):158-62; Schubeler D, Francastel C, Cimbora DM, Reik A, Martin DI, Groudine M. 2000. Nuclear localization and histone

160   acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. Genes Dev. 14(8):940-950.) The implication of these findings is that during cellular differentiation, the nucleus is reorganized in a way that permits the regulation of all relevant genes for a particular cell type. Similarly, there is an implication that disease progression also has an impact on nuclear organization and gene regulation.

165   As a result, leading cytogenetics researchers are attempting to find and understand patterns in chromosome relational arrangements in an ad hoc fashion, developing their own features for use with standard FISH images (Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. J Cell Biol. 145(6):1119-31.; Boyle S, Gilchrist S,

170   Bridger JM, Mahy NL, Ellis JA, Bickmore WA. 2001. The relational organization of human chromosomes within the nuclei of normal and emerin-mutant cells; Bridger JM, Boyle S, Kill IR, Bickmore WA. 2000. Re-modeling of nuclear architecture in quiescent and senescent human fibroblasts. Curr Biol. 10(3):149-52.; Nagele R, Freeman T, McMorrow L, Lee HV. 1995. Precise relational positioning of chromosomes during

175   prometaphase: evidence for chromosomal order. Science. 270(5243):1831-5.; Allison DC, Nestor AL. 1999. Evidence for a relatively random array of human chromosomes on the mitotic ring. J Cell Biol. 145(1):1-14.). For example, Misteli has used nearest neighbor (Parada LA, McQueen PG, Munson PJ, Misteli T. 2002. Conservation of relative chromosome positioning in normal and cancer cells. Curr Biol. 12(19):1692-7.),

180   Allison has used radial angle between chromosomes (Allison DC, Nestor AL. 1999. Evidence for a relatively random array of human chromosomes on the mitotic ring. J Cell Biol. 145(1):1-14) and Bickmore (Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. 1999. Differences in the localization and morphology of chromosomes in

the human nucleus. J Cell Biol. 145(6):1119-31) has used distance from the boundary of

185   the nucleus.

These examples demonstrate that scientists are using application specific image analysis approaches that generate only a few patterns (e.g. Voronoi based regularity index, MST vertex length, nearest neighbor etc.) for analysis. But unfortunately, reliance on a single

190   pattern or small pattern set reduces the chance that important patterns will be detected.

Biological samples such as SKY images of the metaphase rosette (Figure 2A) are non-standard, and often are distorted and arbitrarily oriented. It is difficult to process a large number of SKY images in a robust and accurate fashion that allows result accumulation,

195   which is critical for population analysis to confirm hypotheses or discover subtle pattern differences. The prior art image processing software is unable to handle these large variations. It is impractical to manually normalize the samples. It is highly desirable to have an image analysis tool to automatically normalize the distortion and inter-sample variations among input images for robust and accurate, automated measurements across

200   multiple samples.

A key limitation of current approaches to relational pattern detection is the lack of comprehensive relational pattern configuration features for use in pattern detection. This is due to the tedious nature of manual analysis as well as a lack of image analysis

205   technology. This limitation prevents the detection of subtle differences or higher order (non-trivial) relations within a class or between classes. By using more relational pattern features, users will be more likely to tell the difference between a pattern's signal and noise and could find meaningful patterns efficiently. Also, leading laboratories have indicated that pattern features must be easily understandable so that outcomes can be

210   validated.

Scientists need an integrated image processing and data analysis tool that can enable the creation of many user defined relational pattern features, and support interactive feature

mining for pattern detection and analysis; all in a package that is easy to use and requires
215   no programming. No such tool or combination of tools currently exists.

Objects and Advantages

In a preferred embodiment, the invention includes an advanced analysis tool called a
PatternMap that can be used to detect and analyze strong intrinsic relational patterns
220   within an experimental condition, as well as discriminating relational patterns that can
differentiate experimental conditions. The tool is easy to use and understand, and is
general purpose for use with any biological image application. Requiring only
rudimentary programming, it enables users to efficiently create their own relational
patterns with which to rapidly evaluate the relational arrangement of object classes. It
225   provides an efficient interface for users to rapidly try out different patterns for enhanced
understanding and hypothesis validation, and it supports the comparison of patterns
across experimental conditions.

The primary objective of this invention is to provide a general purpose tool that can be
230   used for the detection and analysis of relational patterns in the relational arrangement of
individual objects belonging to large numbers of object classes in biological images.
Another objective is to normalize the distortion and inter-sample variations among input
images for robust and accurate, automated core measurements across multiple samples.
A third objective is to allow users to develop their own human comprehendible relational
235   pattern features and combined pattern features with which to detect patterns through an
innovative recipe creation interface that requires no programming. The fourth objective
is to provide an innovative pattern detection and analysis framework, called a
PatternMap, that enables the detection and validation of intrinsic patterns within an
experimental condition. The fifth object is to enable the comparison of PatternMaps that
240   illuminates critical differences in patterns between experimental conditions using
PatternMap operations. The sixth objective is to enable genetic diagnosis using
PatternMap or PatternMap derived features.

Summary of the Invention

A method for the detection and analysis of patterns receives an image containing object
245   labels and performs relational feature development using the input image to create at least
one pattern map. It then performs relational feature analysis using the at least one pattern
map to create a relational feature analysis result. The pattern detection and analysis
method further comprises a recipe for automation control and includes determination of a
genetic anomaly.
250

A relational feature development method receives an image containing object labels and
performs core measurement table development using the input image to create at least
one core measurement table. It then performs feature table production using the at least
one core measurement table to create at least one feature table. It also performs
255   PatternMap creation using the at least one feature table to create a PatternMap. The
relational feature development method further comprises a PatternMap integration and
update step to create an updated PatternMap.

A boundary distance measurement receives an image containing object labels and
260   performs structure object mask production using the input image to create structure object
mask. It then performs inner distance transform using the structure object mask to create
inner distance transform image and finds individual object centroid using the input image
to create individual object centroid output. In addition, it finds object boundary distance
using the individual object centroid and the inner distance transform image to create
265   object boundary distance output.

Brief Description of the Drawings

The preferred embodiment and other aspects of the invention will become apparent from
the following detailed description of the invention when read in conjunction with the
accompanying drawings, which are provided for the purpose of describing embodiments
270   of the invention and not for limiting same, in which:
Figure 1A shows the nucleus channel of a standard FISH image of an interphase nucleus;

Figure 1B shows a first labeled gene loci channel of a standard FISH image of the
interphase nucleus used to identify the first specific chromosome;

Figure 1C shows a second labeled gene loci channel of a standard FISH image of the
275   interphase nucleus used to identify the second specific chromosome;

Figure 1D shows the segmented object of Figure 1A;

Figure 1E shows the segmented objects of Figure 1B;

Figure 1F shows the segmented objects of Figure 1C;

Figure 2A shows the CCD image in the SKY image creation;

280   Figure 2B shows the mask image created in the SKY image creation;

Figure 3 shows the processing flow for the PatternMap system;

Figure 4 shows the processing flow for the relational features development method;

Figure 5A shows a rosette mask from a SKY image. It can be used as a reference
structure with which to calculate core measurements;

285   Figure 5B shows a nucleus mask from a standard FISH image. It can be used as a
reference structure with which to calculate core measurements;

Figure 6 shows a rosette structure object can be produced from an image of individual
object masks through a morphological closing operation;

Figure 7A shows an inner distance transform images produced by performing an IDT
290   operation on the rosette mask image of Figure 5A;

Figure 7B shows an inner distance transform images produced by performing an IDT
operation on the nucleus mask image of Figure 5B;

Figure 7C shows an inner distance transform images produced by performing an IDT
operation on the entire image;

295   Figure 8 shows individual object masks overlain on the structure mask IDT image of
Figure 7B;

Figure 9 shows an ODT operation is performed on the chromosome mask generated from
the union of Figure 1E and Figure 1F;

Figure 10A shows an input object mask image;

300   Figure 10B shows the unweighted ZOI partition of Figure 10A;

Figure 10C shows an ODT weighted ZOI partition of Figure 10A;

Figure 11A shows the masks of two objects i and j;

Figure 11B shows overlay of the mask of object i onto object j's ODT image;

Figure 12A shows a PatternMap visualization example;

305 Figure 12B shows a plot of the data corresponding to the horizontal region of the PatternMap elements in Figure 12A;

Figure 12C shows a plot of the data corresponding to the diagonal region of the PatternMap elements in Figure 12A;

Figure 12D shows sub-feature histograms for a given element of Figure 12A;

310 Figure 12E shows image montage for a given element or element sub-features of Figure 12A.

Detailed Description of the Invention

I. Overview

315 In one embodiment, the PatternMap of this invention is a general purpose method for the detection and analysis of automatically generated or user defined patterns in images containing large numbers of object classes with variable number of objects belonging to each class. Given input image(s) containing object labels such as N number of segmented objects belonging to C number of object classes, the PatternMap system is composed of

320 two steps as shown in Figure 3; relational features development 308, and relational pattern analysis 312. The relational feature development method uses the object labels 302 containing in the input image 300 to create at least one pattern map output 304. The relational pattern analysis step 312 uses the at least one pattern map 304 to create a relational pattern analysis result output 306. In addition, the relational pattern analysis

325 step 312 could also determine genetic anomaly result 320. The relational pattern analysis 312 is composed of PatternMap viewing 314 and PatternMap operations 316. The PatternMap development interfaces can be used to create user-defined or automatically generated recipes to support automation 310.

330 As shown in Figure 4, the relational features development method is composed of core measurement table development 406, feature table production 408, PatternMap creation 410 and PatternMap integration and update 412. The core measurement table development step 406 uses the input image 300 to create at least one core measurement

table output 400. The feature table production step 408 uses the at least one core

335   measurement table 400 to create at least one feature table output 402. The PatternMap

creation step 410 uses the at least one feature table 402 to create a PatternMap output

304. The PatternMap integration and update step 412 integrates a PatternMap 304 into

an existing PatternMap 414 through an update process to create an updated PatternMap

404.

340

Relational Pattern Features Development

In one embodiment, the invention consists of a set of core image measurements that can

be manipulated by the user or by computer program to generate a nearly unlimited

345   number of features to describe the relational arrangements of objects in a biological

image. These measurements are generated automatically using image processing

technologies disclosed by Lee, S.J.J. and Oh, S., entitled "Intelligent Relational

Reasoning" United States patent application number 10/411,437 filed April 9, 2003, and

a second patent application by Lee, S.J.J. and Phan, T. entitled "Method for adaptive

350   image region partition and morphological processing", United States patent application

number 10/767,530 filed January 26, 2004, which are incorporated in their entirety

herein. The PatternMap recipe creation interface allows computer automatic generation

of features or manual feature development by users or both to develop features and

feature combinations with which to detect patterns from these core measurements

355   requiring little or no programming.

Relational Pattern Analysis

Once the user directed or automatic relational pattern feature or features combination has

360   been merged into a user defined pattern, the system can automate the production of a

PatternMap that enables the detection and validation of intrinsic patterns within a cell

type. The PatternMap consists of a novel data structure, processing operations and an

easy user interface designed for the purpose of visualizing C x C object class

relationships across an image set. Each row of the PatternMap is essentially a distribution

365   of object classes relations to the C other classes (including itself). In an embodiment, the
       PatternMap include many visualization tools and multi-level object linking that can
       support PatternMap interactive result visualization, processing modification, and outcome
       validation.

370   Typically an image set corresponds to an experimental condition. With the PatternMap it
       is possible to create multiple PatternMaps, each one for a given experimental condition
       (e.g. disease, test, control, etc). PatternMap operations enable the comparison of
       PatternMaps representing different object classes. Differences between PatternMaps can
       be easily detected through inter-map arithmetic and logical operations, as well as with
375   automatic inter-map statistics that highlight differences.

       II. Inputs

       Inputs to the PatternMap system are mask image(s) of N number of segmented objects
380   belonging to C number of object classes. Object segmentation can be performed using
       standard image segmentation and classification tools on biological images. Objects could
       be segmented from standard fluorescence microscopy images and FISH images, as well
       as next generation spectral imaging, emission fingerprinting, spectral karyotyping or
       MFISH images. Those skilled in the art will recognize that the invention can just as easily
385   be extended to address diagnostic images such as digital X-ray, ultrasound, computed
       tomography, magnetic resonance imaging, positron emission tomography and of course
       standard CCD images produced from surgical procedures involving endoscopes. Also,
       the pattern could be derived from not only relational relations but also spectral, temporal,
       or other relations.
390
       III. Relational Pattern Features Development

       III.1 Core Measurement Table Production

395   In one embodiment of the invention, a set of core measurements from which the user or a

computer can assemble nearly any combination of relational patterns for analysis is in the PatternMap. The core measurement (CM) table stores the core measurement data through automatic or semi-automatic processing on the mask images. These tables are all that is needed for follow on analysis so that the raw images do not have to be reprocessed. In the

400   manual mode, a user interacts with the core measurement tables through a graphical user interface (GUI) to create relational pattern features of their choice.

In one embodiment of the invention, core measurement tables include *conditional tables* and *relational tables*. *Conditional tables* are one-dimensional tables (0-N) or other

405   similar sizes that store a core measurement for an individual object. In one embodiment of implementation, N is defined as $N = \sum_{0}^{C} \sum_{0}^{B} 1$ where C is the number of object classes and B is the (variable) number of objects for a given class. They are called conditional tables because they are used to condition the relational tables by filtering out entries that don't meet a defined condition based on the measurement in the conditional table. In one

410   embodiment of the invention, the measurements include but are not limited to:

- Boundary distance: shortest distance (in pixels) from each object to a structural boundary. The boundary could be one of the object masks, such as the nuclear mask, or rosette mask shown in Figure 5A. The boundary could also be the image

415   frame.

- Radial distance: the radial distance is the angular distance (in radians) of an individual object's centroid vector from the horizontal axis, defined relative to the boundary determined by the user.

420   Relational tables are N x N tables or other similar sizes where each element of the table encodes a core measurement describing a relation between two objects. In one embodiment of the invention, the measurements include but are not limited to:

- Object distance: the lowest number of individual objects between two individual objects; the unit is in number of objects (the distance between adjacent objects is

425   zero)

- Radial difference: the difference in radians between two objects.
- ☐boundary difference: the normalized absolute difference in the boundary distance of two objects.
- Pixel distance: the shortest distance between two objects.

430

III.1.1 Boundary Distance Table Production

The boundary distance table is a one-dimensional table where the shortest distance from each individual object centroid to the structure boundary is stored. The boundary

435   distance measurement performs structure object mask production using the input image to create a structure object mask. It then performs an inner distance transform using the structure object mask to create an inner distance transform image. It also finds individual object centroids using the input image to create individual object centroids. It then finds object boundary distance using the individual object centroid and the inner distance

440   transform image to create the object boundary distance output.

III.1.1.1 Structure Object Mask Production

The first step creates the structure object mask. As mentioned above, the structured object

445   could be created from one of the object classes, such as the nuclear envelope 502 in Figure 5B, or from some structure in the image, such as the chromosome rosette 500 shown in Figure 5A, or the image boundary. In some cases, like the nuclear mask for instance, no processing is needed, but in others some image processing is required to create the structure object mask, as shown in Figure 6 where a morphological closing

450   operation is performed to create a rosette structure object mask 500 from an image containing individual chromosome masks 600.

III.1.1.2 Inner Distance Transform (IDT) Image

455   Inner distance transformation creates a distance image that assigns the shortest distance from each of the mask object pixels to a background pixel (Meijster A, Roerdink BTM,

Hesselink WH. 2000. A general algorithm for computing distance transforms in linear time; Kolountzakis MN, Kutulakos KN. 1992. Fast computation of the Euclidean distance map for binary images. Information Processing Letters). Using the structure

460   object mask image, an inner distance operation is performed on the object mask. The inner distance transform creates a transformed image where for every pixel inside the reference structure, the shortest integer pixel distance from that location to the structure object boundary is encoded as that pixel's value. Thus, pixels outside or on the structure boundary are assigned a zero, pixels one pixel inside the structure boundary are assigned

465   a one, pixels two pixels inside the structure boundary are assigned a two and so forth such that the highest number will be encoded in pixels that are located the farthest distance from the structure boundary. Those skilled in the art should recognize that different distance metric could be used. Euclidian distance (norm 2) is used in Figure 7.

470   Figure 7A shows an IDT image for the rosette structure shown in Figure 5A. Figure 7B shows an IDT image for the nucleus mask image shown in Figure 5B. Figure 7C shows an inner distance transform images produced by performing an IDT operation on the entire image. This is applicable in the case where the image frame is the reference structure.

475

III.1.1.3 Find Individual Object Centroid

Given an image composed of individual object masks, each object's centroid can be computed from the 1st moment of each component. For a given mask R, the centroid of R

480   (x_0, y_0) can be found as:

$$x_0 = \frac{\sum\limits_{(x_i,y_i)\in R} x_i}{\sum\limits_{(x_i,y_i)\in R} 1} \cdot y_0 = \frac{\sum\limits_{(x_i,y_i)\in R} y_i}{\sum\limits_{(x_i,y_i)\in R} 1}$$

III.1.1.4 Find Object Boundary Distance

485

For each individual object i in the boundary distance table, the boundary distance of i is the value in the pixel location in the IDT image that corresponds to chromosome i's centroid location in the mask image. Figure 8 shows the masks of the individual chromosomes from Figure 1 overlain on the nuclear structure mask IDT image. Each

490   individual chromosome boundary distance can be found by the pixel value in the IDT image at the location corresponding to the chromosome mask centroid. Note that if an object is outside the reference structure object, its boundary distance will naturally be assigned to zero. Those skilled in the art should recognize that other methods for finding object boundary distance could be used. For example, object boundary distance could be

495   defined as the maximum IDT value within the object mask.

III.1.2 Radial Distance Table Production

The radial distance table is the angular distance (in radians) of each individual mask

500   object's centroid vector from the horizontal axis.

In one embodiment of the invention, the center of the axis is set at the centroid of the structure object mask as described above in III.1.1.1. A meaningful reference object should be chosen, as it provides a reference or coordinate system within which the

505   individual object's relative positions are stored in the radial distance (and boundary distance) tables. Examples of meaningful structures include the chromosome rosette, nuclear envelope, tissue structure such as a colon crypt (Cooper HS, Murthy SNS, Shah RS, Sedergran DJ. 1993. Clinicopathologic Study of Dextran Sulfate Sodium Experimental Murine Colitis. Lab Invest. 69(2):238-49) or if there is no appropriate

510   reference structure in the field of view, the image frame can be chosen. Calculate the centroid for the structure object mask from the 1th moment of the mask, the same as is done above for the individual objects in III.1.1.3.

Set the zero radians horizontal axis (relative to image frame) from the centroid of the

515   structure mask to the edge of the image. For each individual object 0 – N, we measure the

radial displacement of its centroid from 0 and put that value ($0 - 2\pi$ radians) into the Radial Distance table.

III.1.3 Object Distance Table Production

520

The object distance table is an N x N relational table where each table element i x j stores the object distance between individual objects i and j.

Object distance is the fewest number of individual objects between 2 individual objects i

525 and j; the distance unit is in object count. For example, given four individual objects in a line A, B , C, D, E:

- The object distance of (A,B): OD(A,B) = 0
- OD(A,C) = 1 since there is B between them
- OD(A,D) = 2 since there are B and C between them

530 - OD(A,E) = 3 since there are B, C and D between them

III.1.3.1 Adaptive Zone of Influence (AZOI)

AZOI is a zone of influence (ZOI) operation where a component characteristic dependent

535 distance metric is applied. ZOI operation assigns each pixel of the image a unique label of the connected component to which it is closest (influenced). ZOI labeling performs fast region partitioning (Ogniewicz RL, Ilg M. 1992. Voronoi skeletons: theory and applications. IEEE Comp. Vision and Pattern Rec. 63-69; Ogniewicz RL, Kubler O. 1995. Hierarchic Voronoi skeletons. Pattern Recognition nr. 28; 343-359).

540

To produce the object distance table, first we perform an AZOI to find AZOI boundaries having outer distance values less than or equal to a given threshold to find the meaningful transitions between adjacent objects.

545 AZOI provides a method for adaptive image region partition and morphologic processing. The partition could be computed sequentially by a two-pass procedure

similar to distance transformation. In addition to the shortest distance, the method tracks, propagates, and updates the component label where a pixel's shortest distance is derived.

550   The AZOI partition is done for each individual object mask in the input mask image(s). The purpose of the AZOI is to create a non-ambiguous representation across which adjacency transitions can be calculated. In many biological imaging applications, it is desirable for adjacency to be calculated for two objects even if they are not touching. On the other hand, it is often undesirable for two objects to be calculated as adjacent if

555   they are separated by a large expanse of empty space. To address this issue, this embodiment utilizes component outer distance transformation based threshold as a distance metric to constrain the ZOI region partition.

An ODT image is the opposite of the IDT image described above. As shown in Figure 9,

560   given an input image composed of individual mask objects 900, for each pixel located outside the mask objects, the ODT encodes the shortest distance from that pixel to an object mask boundary as that pixel's value 902.

The ODT image can be used to threshold the ZOI partition by removing ZOI

565   boundaries with large ODT values. This enables transition identification that ignores adjacency across large, empty areas. This ODT based threshold can be set by the user or determined automatically. Figure 10A shows an input mask object image from a SKY platform. The non-weighted ZOI image for the individual objects in Figure 10A is shown in Figure 10B, and the ODT weighted ZOI image is shown in Figure 10C. For this

570   embodiment, determine the adjacent pairs from the object labels corresponding to each of the retained ZOI boundaries.

III.1.3.2 Populate the Object Distance Table

575   First we populate an intermediate, object adjacency table by placing each individual's adjacent neighbors (look up from the adjacent pairs) into it's adjacency set (each member Adj_{ij} is adjacent to object i as follows:

$$i = 1 \mid Adj_{11}, Adj_{12}, Adj_{13} \ldots Adj_{1k}$$

580
$$2 \mid Adj_{21}, Adj_{22}, Adj_{23} \ldots Adj_{2l}$$

$$\ldots$$

$$N \mid Adj_{N1}, Adj_{N2}, Adj_{N3} \ldots Adj_{Nm}$$

Next we use this adjacency set information to populate the Object Distance Table, which is a relational N by N table containing the object distance of each object interaction i x j.

585

Conceptually, the algorithm for determining object distance is:

    (1) For a given chromosome i find all its adjacent neighbors

        (a) Label the distance from i to these neighbors as 0

    (2) Group i and its neighbors into group A

590
    (3) Find all of group A's adjacent neighbors

        (a) Label the distance from i to these neighbors as 1

    (4) Repeat and increment distance until there are no more objects

In pseudo code:

595
```
ADJ(C_i)   ∀ i ∈ {1 .... N)

For (i = 1; i ++; i ≤ N)
{
A = ∅; B = ADJ(C_i); d = 0;
```
600
```
while (B ≠ ∅)
{
for (j = 0; j ++; j ≤ N)
{
if(c_j ∈ B && j ≠ i)
```
605
```
dist (ci,cj) = d;  // dist() loads the object distance table
}
d++;
A = A ∪ B;
```

$$B = \cup_{Ck\in A} ADJ(C_k) - A;$$

610   }

           }

III.1.4 Radial Difference Table Production

615   The radial difference table is an N x N relational CM table that for any given individual object interaction i x j the radial difference between object i and j will be recorded.

       a)   Radial difference is measured in radians

       b)   Radial difference is the smallest angle between i and j ($0 - \pi$)

620   c)   Radial difference can be calculated from the conditional radial distance table

III.1.5 Δ Boundary Distance Table Production

Δ boundary distance is an N x N table where for a given object relation i x j, Δ boundary

625   distance is the normalized absolute difference in the boundary distances of each object. It indicates the "In/Out-ness" of the i, j relation. A high score means that one of the pair is near the center and the other is near the boundary. A low score means that both are close together relative to the center or the boundary.

630   Δ Boundary Distance can be derived from the conditional boundary distance table (BD). For a given interaction i x j, Δ Boundary distance is the absolute difference between i and j's boundary distance, divided by the maximum boundary distance in the reference structure object mask.

$$\Delta \text{ Boundary Distance} = \frac{|BD[i] - BD[j]|}{\max_{k}(BD[k])} \quad \text{where BD[i] is the boundary distance value}$$

635   for object i and k is the max entry in the BD array.

III.1.6 Pixel Distance

In one embodiment of the invention, there are two kinds of pixel distance tables; pixel
640   distance average and pixel distance edge. Those skilled in the art should recognize that other methods of pixel distance such as median, minimum, or other statistics could be used. They are all within the scope of this invention.

III.1.6.1 Pixel distance average table: Distance_avg

645

The pixel distance_avg table is a N X N table that for any given individual object interaction i x j records the average of the shortest distance of each pixel in i to the boundary of j.

650   Given an object j 1100, we create the ODT image (Figure 11B) from the boundary of j 1100 as described above in III.1.3.1. We then overlay the mask of object i 1100 onto this ODT image (Figure 11B), and take the average ODT value of the pixels within i from the ODT image as shown in Figure 11A and Figure 11B.

655    $$\text{Pixel Distance\_avg[i][j]} = \frac{\sum_{(x_k,y_k)\in i} ODT(x_k,y_k)}{\sum_{(x_k,y_k)\in i} 1}$$

III.1.6.2 Pixel distance edge table: Distance_edge

The pixel distance_edge table is a N X N table that for any given individual object
660   interaction i x j records the shortest distance between the boundary of mask i to the boundary of mask of j.

Given an object j, we create the ODT image from the boundary of j. Next we overlay the mask of object i onto this ODT image, and take the minimum ODT value along the
665   boundary of i's mask and store it in the Pixel Distance_edge table.

III.2 Feature Table Production

This step integrates the individual objects into object classes and generates multiple C
670   x C feature tables (where C is the number of classes). In most biological applications the
       number of classes is known a priori whereas the number of individual objects is not. In
       the SKY example from Figure 2, there are 20 classes corresponding to the chromosome
       classes of a mouse. One of the innovations and benefits of this invention is enabling users
       to integrate data describing an unknown number of individual object interactions into a
675   more easy to use framework where those individual interactions become part of the
       statistic describing an biological class interaction; which is a more meaningful unit of
       analysis.

       This invention allows users to specify object integration rules through the GUI. The
680   specification could also be performed automatically. These rules involve four steps;
       selecting the input table from the relational CM tables, selecting the feature rule,
       selecting the data treatment, and selecting the class member integration rule. It can be
       repeated multiple times to create multiple feature tables.

685   III.2.1 Select the Input Table

       The features of interest can only be generated using the core measurements from the
       relational CM tables (not the conditional CM tables). A user selects from the object
       distance table, pixel distance_avg table, pixel distance_edge table, radial difference table
690   or Δboundary distance table.

       III.2.2 Select the Feature Rule

       Every element in the input table is subject to a user defined feature rule. The rule acts
695   as a filter that passes data only from object relations that past the rule or test. The output
       is a filtered relationship table. If the value in the element passes the rule, either the value
       or a '1' will be passed to the filtered relationship table (depending on the user selected
       data treatment, see III.2.3 below). If the value in the table element fails the rule, some

value (such as −1, or 0) will be passed to the filtered relationship table indicating that the
700   relation failed the rule.

Rules include element based rules and row based rules. Element based rules include
conditional CM table rules and relational CM table rules.

705   III.2.2.1 Element (E) Rules

Element rules are evaluated on each table element individually, and are measured for
the interpretation of each individual element. In one embodiment, the element (E) rules
include conditional CM table rules and relational CM table rules.
710

III.2.2.1.1 Conditional CM Table Rules

Conditional CM table based rules condition the input relational CM table by using the
one-dimensional conditional tables, Boundary Distance (BD) and Radial Distance (RD),
715   to filter out the horizontal and vertical rows corresponding to the individual objects that
fail the rule. In one embodiment of the invention, these rules include:

a) Boundary Distance (BD) $> \alpha\%$

   i) $(\alpha, \beta$ is user selected percentage of the max BD in the reference structure

720   object mask IDT)

d) $BD < \alpha\%$

e) $\alpha\% < BD < \beta\%$

f) $BD = \alpha\%$

g) Radial Distance (RD) $> \alpha\pi$ radians

725   i) Here α is a user selected number between 0.00 and 2.00

h) $RD < \alpha\pi$

i) $\beta\pi < RD < \alpha\pi$

j) $RD = \alpha\pi$

730   III.2.2.1.2 Relational CM Table Rules

Relational CM table based rules filter the input relational CM table using the element values themselves to create the filtered relationship table.

     a)   Element value (E) > α

735   i) α,β is user selected value, E is the value of the element in the input table for any given object interaction i, j

     b)   E < α

     c) $\beta < E < \alpha$

     d)   E = α

740

III.2.2.2 Row Rules

Row rules are evaluated on each element, but are measured for the interpretation of the row

745

     a)   RowRank > α

         i) α,β is a user selected value between 0 and N

         ii)   Row elements are ranked by element value from smallest to largest (N max)

         iii) Row rank is same for equivalent values (more than one element could have

750   the same rank)

     k)   RowRank < α

     l)   RowRank = α

     m) β < RowRank < α

     n)   RowRank = RowMax - α

755   i)   Here α refers to element rank, not element value

     o)   RowRank < RowMax - α

III.2.3 Choose the Data Treatment

760   III.2.3.1 Pass / Fail

If the table element passes the condition, the corresponding element location in the filtered relationship table gets a 1, if it fails it gets a value indicating failure (e.g. 0 or –1).

765   III.2.3.2 Value

If the table element passes the condition, the element value in the input table is passed to the corresponding element location in the filtered relationship table. If it fails the condition a value indicating failure (e.g. 0 or –1) is passed to the filtered relationship

770   table.

III.2.4 Selecting the Class Member Integration Rule

The class member integration rule is defined by the user and used to create a C x C

775   class based feature table from the N x N filtered relationship table.

It is desirable to create the feature table that for every element i x j, stores a statistic describing the relationship between class i and j in the input image. This invention allows the user to specify a simple class member integration rule to create a statistic describing class i and j's interaction in the input image based on measurements of the individual

780   object interactions of the objects belonging to each class; no matter how many there are.

For example, in the SKY image example of Figure 2, where for a given class interaction i x j, each class i,j has two individual objects A,B, the class member integration rule

785   provides a function for mapping 4 elements in the N x N filtered relationship table to one element in the C x C feature table:

$$FT[i][j] = f(CMF[iA][jA],CMF[iA][jB],CMF[iB][jA],CMF[iB][jB])$$

Where FT = feature table

790   CMF = core measurement feature from the filtered relationship table

Integration rule functions include:

795   (1) Min

                      (2) Rank 2

                      (3) Rank 3

                      (4) Sum

                      (5) Avg

800   (6) +,-,*,/

III.3 PatternMap Creation

Assuming multiple feature tables were created, this step allows the user or automatic

805   module to define the method for combining the feature tables together into a single

pattern map; the tables are combined cell-to-cell. The possible ways to combine the tables

include simple arithmetic, logical operations, min, max, weighted sum, etc. It is also

possible to perform scalar operations on the CP tables. Scalar operations include plus,

minus, multiple, integer divide, floating point divide and mod. Thus, the inputs are the

810   feature tables, scalar values and operations specification, and the output is a single image

pattern map (PM). This step can also be repeated to generate multiple pattern maps.

III.4 PatternMap Integration and Update

815   In this step the user defines rules for integrating the single image PMs into a PM for

the entire data set (called image set PM). There are multiple 'types' of PatternMaps that

are distinguished by the feature rules that create the elements' statistic. For example, one

single image PM could be a 3 Nearest Neighbor accumulation PM (an accumulation is a

weighted integration of multiple pass/fail feature tables), and another could be a 'Distal-

820   ness' pass/fail PM.

PM update rules are able to integrate both the same and different types of PMs. Therefore, the PM update rule takes in an arbitrary number of PMs and scalar values and combines them into a single image set PM. As such, it is identical to the singe image PM creation rule of III.3 above. The difference here is that an additional step is added to define how the merged PM is updated into the PM for the image set.

This step can also be repeated to generate multiple image set PatternMaps.

In one embodiment of the invention, the user or automatic module selects rules from the following list:

    i)   PM integration rule (algebraic expressions) for single image PMs

        (1) +, -, *, / (integer divide), ÷ (floating point divide), % (mod)

        (2) AND, OR, XOR, min, max

        (3) A potential expression could look like: PM3 = PM1 – PM2

    ii) PM update rule for accumulating integrated PMs into the data set PM

        (1) Min

        (2) Max

        (3) Sum

        (4) Average

IV. Relational Pattern Analysis

IV.1 PatternMap Viewing

The PM consists of a data structure, processing operations, and an easy to use interface designed for the purpose of visualizing C x C object class relationships across an image set. As shown in Figure 12A, a PM is a C x C matrix where each axis contains the complete class set 1 - C. At each matrix intersection or element, the pattern relationship of a given element i x j can be reviewed, including class membership relations at element i x i. Each element of the PM contains a score (summary statistic)

directly related to the likelihood of the presence of a user-defined pattern derived from the accumulation of all the images for the cell type. As shown in Figure 12A, map elements can be shaded or color coded to bring attention to strong scores. The shading or

855   color coding can be manually adjusted with the threshold bar on the left side of the PM (Figure 12A). Each row of the PM is essentially a distribution of a class' relations to the C other classes (including its own class members). Bar charts of this distribution can be displayed for any row as shown in Figure 12B. Class member statistics (shown here as the diagonal of the pattern map in Figure 12C) can also be viewed as bar charts. As

860   shown in Figure 12D and Figure 12E, images and sub-feature data can be explored for a given element. The sub-feature data corresponds to the input feature tables. Figure 12D shows the histogram of the input data for a given PM element broken down by sub-feature; in this example there are three sub-features. Images for an entire element, or only a specified element sub-feature can be presented in a montage for user review (Figure

865   12E). The PM can be incrementally updated. Therefore, data from new images could be easily added into an existing PM.

IV.2 PatternMap Operations

870   PM operations allow transformation and information extraction of a PM. For example, an entropy measurement and other distribution statistics to help detect the unevenness of the distribution (that is, significant patterns) can be performed by PM operations. Users can benefit from many views of the data either by summary statistics or data visualization.

875

PM operations are used primarily to compare across PMs. Inter-map logic and arithmetic operations can be performed and a result PM can be viewed. For example, one PM corresponding to a cell line could be subtracted from another PM of a differing cell line, the difference PM will show the relational patterns that differ between cell lines.

880

All of the charting of a single image set PM can be done to compare the same statistics for multiple image set PMs. For example, a row distribution bar chart can be shown with multiple bars per class, where each bar corresponds to a different image set PM.

885   Those skilled in the art should recognize that many matrix operations and statistics could be used for PM operations.

V. Recipes

890   The recipes support automation of the above steps and cover everything from image loading to PM operations. Recipes can be used to automate the processing of image samples into an image set PM via PM update rules, or even to process images and store results to a database or file without ever reviewing the intermediate processes.

895   V.1 PatternMap Recipe

The PM recipe contains an image loading recipe, a feature table production recipe, a PM creation recipe, PM update recipe, a PM operations recipe, and an output recipe.

900   In one embodiment of the invention, the user begins by selecting 'create recipe' from **Manage > CreateRecipe**, and chooses **PatternMap.** Next the user gives a name to the PM.

This brings up the **Design Recipe** window, which is an icon based interface for
905   programming recipes. The **Design Recipe** window is specific to this PM recipe being created. The user can choose to work in the **Design Recipe** window, or select interactive mode that operates in the normal environment.

The user does not have to have a complete sequence of recipe steps. For instance,

910   given 50 input images the user could automate the creation of the image set PM without automating PM operations or data output. The user could automate the production 50*N chromosome pair tables, which is allowed with a warning.

Once the entire recipe path has been configured and each step saved, the user can
915   choose to save and close or execute the PM recipe.

V.2 Image Loading Recipe

The user drags an image loading recipe from the tray in the design window and
920   double clicks on it. This brings up a dialogue in which the user can specify the file folder or database from which the input images are to be found.

V.3 Feature Table Production Recipe

925   Next the user drags a feature table production recipe icon from the tray onto the canvas and uses a connector to show that this is the next step after image loading. The user double clicks on the icon and brings up the chromosome integration rule interface, which is the same regardless of whether one is in interactive mode or in the **Design Recipe** window. The GUI has a 'Set Recipe' button that will assign the selected options
930   to the icon in the Design Window.

If the user is in interactive mode, and accesses this GUI via the normal software environment, when he/she clicks on 'Set Recipe', a feature table production recipe icon will be automatically added to the canvas in the **Design Recipe** window.
935

V.4 PatternMap Creation Recipe

Same as in step V.3 above, except that the interface and icon are specific to PM creation.
940

V.5 PatternMap Update and Integration Recipe

Same as V.3 and V.4 above, except that the interface and icon are specific to PM update and integration.

945

V.6 PatternMap Operations Recipe

Same as V.3, V.4 and V.5 except that the interface and icon are specific to PM operations recipe.

950

V.7 Output Recipe

This is accessed in the **Design Recipe** window, the user can select which data to output, and how to output it. PM result output can of course be done in many formats

955   including text file, Microsoft Excel file, XML and database formats

The invention has been described herein in considerable detail in order to comply with the Patent Statutes and to provide those skilled in the art with the information needed to apply the novel principles and to construct and use such specialized components as are

960   required. However, it is to be understood that the inventions can be carried out by specifically different equipment and devices, and that various modifications, both as to the equipment details and operating procedures, can be accomplished without departing from the scope of the invention itself.